Van de Walle, S. (2016). The experimental turn in public management: How methodological preferences drive substantive choices. In O. James, S. Jilke & O. Van Ryzin (Eds.), *Experiments in public management research*. Cambridge: Cambridge University Press.

THE EXPERIMENTAL TURN IN PUBLIC MANAGEMENT: HOW METHODOLOGICAL PREFERENCES DRIVE SUBSTANTIVE CHOICES

Abstract

In their search for empirical credibility, public management researchers may search for topics that lend themselves more easily to experimental methods. This chapter looks at how the use of the experimental method may change the questions asked in the field of public management. Possible consequences are a focus on discrete interventions and marginal changes, and a shift away from studying public organisations themselves to a study of the behavior of individuals within these organisations. The chapter shows how methodological preferences may drive substantive research choices. The chapter first discusses different reasons why experiments have become popular in public administration, including a search for credibility, fashion, and custom. It then shows how the choice of an experimental approach influences what topics or questions are studied, and how they are studied. Subsequently, a number of implications for the future of experimental research in the field of public management are outlined including the need for enhancing experimental realism, ethical challenges, replication needs, a possible move to formal modelling, and changes in publication practice.

Keywords: RCT, field experiment, credibility, experiment, external validity, realism

THE EXPERIMENTAL TURN IN PUBLIC MANAGEMENT: HOW METHODOLOGICAL PREFERENCES DRIVE SUBSTANTIVE CHOICES

In their search for empirical credibility, public management researchers may search for topics that lend themselves more easily to experimental methods. That in itself is not surprising in a field that has only recently 'discovered' experimental research. This chapter looks at how the use of the experimental methods may change the questions asked in the field of public administration and management. Possible consequences are a focus on discrete interventions and marginal changes, rather than on protracted system changes, and a continuing shift away from studying public organisations themselves to a study of the behavior of individuals within these organisations.

Each change in methods and methodological preferences brings with it a change in what is studied. Sometimes, this is because new methods allow for studying things and phenomena that could not be studied in the past. In other cases, changed methodological preferences themselves change the researcher's focus. An example of the first is the (belated) adoption of multilevel statistics in public management research, which stimulated a shift towards studying 'large-N' multilevel settings. Hence the interest in studying local governments, schools or hospitals because such setting require using more advanced multilevel designs. Another example is the renewed interest in political and public communication at a time when tools for analysing social media data became widespread. An example of the latter is the quantification of the discipline and the related use of (large) surveys, which has resulted in a shift to individuals (and their attitudes and behaviours) and away from the field's traditional focus on institutions.

In this chapter, I show how methodological preferences may drive substantive research choices. I first discuss different reasons why experiments have become popular in public administration.

I then show how the choice of an experimental approach influences what topics or questions are studied, and how they are studied. I end by sketching a number of implications for the future of experimental research in the field of public management and for publication practice.

1. Substance over method, or method over substance?

The choice of research tools and designs depends on the substance studied in a discipline. For a historian, archival work obviously makes more sense than setting up an experiment, which requires a prospective rather retrospective approach. As a rule, studying a topic is best done using a variety of methods. Methods however may come to dominate substance because, as Kinder (2011) explains, 'Relying exclusively on one method constricts the range of questions that seem worth pursuing. Which questions are interesting and which are not is seen through the filter of what one is able to do. Methodological preoccupations inevitably and insidiously shape substantive agendas' (Kinder, 2011: 526). Methodological choices and preferences may drive what questions are seen as worth studying. Concerns with either external or internal validity likewise influence the choices that are made in doing research and selecting a topic or cases. Concerns in the discipline with external validity rather than internal validity lead to a focus on ever larger surveys, an increasing focus on comparative case methods, and a rise in cross-sectional and cross-national research. Single-organisation studies and case studies have lost much of their standing, despite the fact that such studies have brought the discipline some of its seminal theoretical insights (just think about gypsum plants (Gouldner, 1964) and forest rangers (Kaufman, 1960). Serious concerns about internal validity coupled with a desire to strengthen the discipline's empirical credibility have now stimulated a move to experimental methods. In line with Kinder, I argue that diversification in the methodological repertoire is essential to understand important questions. Also in line with Kinder (2011) and Kinder and Palfrey (1993), I argue that experiments are a great addition to the field of public management.

2. Why experiments?

Public management as a field clearly had some catching up to do when it comes to using experiments (and indeed methods in general), especially with other disciplines like political science and economics pecking at its borders (see Chapter x for a review of experiments in public management). In development studies and development economics, the randomized controlled trial (RCT) revolution included many experiments involving public service delivery in the education or health sector (see Banerjee & Duflo, 2012, for examples). Scholarship on decision-making in public management has hardly moved to using experiments, while it has in other disciplines. Political scientists and economists increasingly move onto PA's 'turf' when conducting experiments involving all kinds of public decision-making or the delivery of public services and public goods. How can we explain the discipline's belated turn to experiments as a research method? In this section, I distinguish between three potential explanations: fashions and bandwagon effects, the search for credibility, and custom.

2.1.Experiments as a fashion?

In his Presidential Address to the American Sociological Association in 1975, Lewis Coser talked about a 'method in search of substance', reflecting about editors' decisions to stop accepting papers not using certain -- then innovative -- quantitative methods. In other words, newness in methods appeared to be seen as a quality criterion. Indeed, for the ambitious researcher, it makes sense to surf the wave of attention for experimental methods: journals and special issues court your copy and publishers want your handbook. Such academic bandwagons are a regular occurrence in most disciplines. Cass Sunstein, in this respect, referred to cascade effects: 'Academics, like everyone else, are subject to cascade effects. They start, join, and accelerate bandwagons. More particularly, they are subject to the informational signals sent by the acts and statements of others. They participate in creating the very signals to which they respond. Academics, like everyone else, are also susceptible to the reputational pressures

imposed by the (perceived) beliefs of others. They respond to these pressures, and by so doing, they help to amplify them. It is for these reasons that fads, fashions, and bandwagon effects can be found in academia' (Sunstein, 2001: 1251).

He links this, amongst other factors, to utility functions, such as entry to the job market or access to journals. The proliferation of special issues and calls for papers with a focus on experimental public administration and public management is a good stimulus for scholars to move into experimental research. Being among the first to use a method that is gaining recognition gives one advantages on the market for jobs, attention and resources, where novelty and newness are important rhetorical devices.

Cascades may be information-induced. Especially in a field such as public management, where attention to methodology remains, despite many improvements, still relatively limited and with a lot of uncertainty, the work of well-known scholars and topical choices made by leading journals send information signals to the wider community, creating a cascade of experiments. Something like this appears to happening now in public management research, as several leading journals have published special issues on experiments and a number of well-known scholars have jumped on the experimental research bandwagon. As the review of the published literature in Chapter X documents, the number of experiments published in public administration and management journals has grown rapidly in the last few years.

2.2. The search for credibility

Public management has for a long time been regarded as merely applied, and has borrowed many of its theories and methods from more established disciplines such as political science, sociology or economics. The focus on experiments fits within a wider trend in public management research of attempting to be seen as 'academic' rather than as an applied discipline (Gadellaa, Curry & Van de Walle, 2015). Such a transformation means adoption of

the tools or at least the artefacts of real scientists. This transformation is helped by the simultaneous move to a more widespread use of field experiments within governments themselves and of RCTs in applied research. The latter trend has also helped the field not just to strengthen the internal validity of its work, but also the face validity of the work done, because experiments are increasingly being seen as valid ways of understanding the world of policy-making and service delivery.

Empirical credibility is thought to come mainly from quantitative approaches, and from high internal validity. Neighbouring disciplines, such as economics or political science have had their own credibility movements before. Choice of research methods influences the perceived quality of the research. Quantifying provides studies with an aura of precision. Changing the name of one's institute or department to 'laboratory' imprints an aura of seriousness on what is going on inside – it will certainly convey a message that serious research is going on, rather than something 'soft'. An additional advantage of the experimental approach is that the simplicity of the designs allows for a high transparency of the research, something which is highly valued following a series of scandals in the social sciences (some of which, it has to be said, related to experimental research).

2.3.Custom: The law of the instrument

Custom is another mechanism that may explain attention to specific methods. Kaplan, in his 1964 classic 'the conduct of inquiry' called this 'the law of the instrument', which states 'that a scientist formulates problems in a way which requires for their solution just those techniques in which he himself is especially skilled' (Kaplan, 1998(1964): 28). In other words, what the scientist is able to do determines the method used. Methodological myopia is a less complimentary description of this phenomenon. Method preferences are thus not merely lead by substantial questions and the suitability of the method to the research at hand.

This mechanism is one that may work for or against a rapid proliferation of experimental research in the discipline. Most scholars trained in public management departments have received very little training in experimental research. Indeed, using experiments was confined to psychologists and a couple of behavioural economists and political scientists. Younger generations are now just starting to grow up with at least some of the basics of experimental design and analysis in their graduate school training. Still, the absence of experimental training in most public management programmes thus leads to a lack of the use of this method. Other disciplines, such as economics, could profit from the presence of marketing scholars to introduce experimental training into doctoral training curricula, thereby preparing future generations of scholars. Also, in development studies, attention to RCTs has grown substantially (Banerjee & Duflo, 2012). In public management, there is as yet no 'school' known for experiments. To make experimental methods flourish, doctoral students need to be acculturated and training priorities revisited.

3. How experiments change what is studied

Using experimental methods, just like using any other method, requires redesigning research questions in such a way that they are amenable to experimental testing. This means some questions are more suitable for experimental testing than others. Examples are for instance whether incentives work, or whether providing (performance) information alters behaviour. Experimental approaches, in other words, are excellent means to further our knowledge about some parts of the world of public management and administration, but this selectivity in topics may come to the detriment of other parts of the field. McDermott's review for instance showed that the bulk of experimental work in political science concerned voting (McDermott, 2005). When researchers, because of the reasons mentioned in the previous section, select their research question based on preference for using experimental methods, methodological preferences may shape what topics are researched. In a critical account of RCT's by recent

Economics Nobel laureate Angus Deaton, he describes the process as follows: ‘This goes beyond the old story of looking for an object where the light is strong enough to see; rather, we have at least some control over the light but choose to let it fall where it may and then proclaim that whatever it illuminates is what we were looking for all along.’ (Deaton, 2010: 429).

A more specific point beyond the general tendency of users of experimental methods to focus on certain topics is the observation that experimental methods can, in fact, only deal with very specific interventions. Experimental designs are particularly useful to study certain topics and processes, and have shown their value in particular in situations where discrete interventions take place. For this reason, experimental research pays particular attention to borders, sudden changes, or cut-off points, such as competitive grant allocations, entry exams, major policy changes, or redistricting. Because of their nature, experiments are also very well suited to test effects of specific interventions in evaluation studies, or for testing specific clinical treatments. Specific quasi-experimental designs allow for studying such situations. Think for instance about regression discontinuity designs that are used to study the effect of cut-off points, for example in competitions or selection procedures.

Such approaches are obviously not without value, but the real question is: how often do such situations occur in the real world? Interventions in the real world are seldom discrete. A treatment in real life is almost never a discrete event, but always confounded by other factors and preceded by policy debates and announcements. Researchers relying on natural or quasi-experiments have frequently tried to solve this issue by redefining major interventions as if they were discrete. In this way, a major social change (a power blackout, a change in registration systems etc.) is redefined as if it were just a simple decontextualized intervention. Such decontextualisation also makes it easier to identify instrumental variables.

Interventions generally interact to generate outcomes (Pawson & Tilley, 1997). In the real world interventions are often part of a wider package, they are complex, and they are not isolated from their environment. Still, this is something that can be tested using experiments, but it requires very large numbers of experiments, with many variations in treatments and contexts. One could then wonder whether experiments, despite their strong internal validity, are still worth the effort. Experiments are good for establishing causal effects, and that makes them highly effective in situations where causal effects are thought to be fairly simple and probably linear. In other words, they are not particularly useful to unravel complex processes or interactive systems. For the field, this may mean a welcome increase in studies on small changes, marginal improvements, and specific discrete interventions, but a marginalisation of studies on system change. Cartwright, in a paper on the use of RCTs, commented that narrowness of scope may be the price to pay for the obvious benefits of experimental research (Cartwright, 2007: 11). In a survey among criminologists on why experiments were not used, a desire not to simplify complex social processes was mentioned prominently, alongside skills, practical issues, and ethics (Lum & Yang, 2005).

The issue is not necessarily the external validity of the findings, but the external relevance of the stimulus being tested. Deaton gives the argument that in the real policy world, outcomes of a treatment may be different when for instance 'everyone is covered by the treatment rather than just a selected group of experimental subjects'. He gives a concrete example: 'Small development projects that help a few villagers or a few villages may not attract the attention of corrupt public officials because it is not worth their while to undermine or exploit them, yet they would do so as soon as any attempt were made to scale up' (Deaton, 2010: 448). An experiment on how subjects react to different forms of performance information (see, e.g. James, 2010) on public programmes may well depend to a large extent on the prevalence of publicly available performance information or the existence of a tradition of a government that

communicates openly about performance in the society within which the experiment is fielded. It also depends on whether everyone in society consults such information, or just a small group of people. We thus don't know whether the same effect would be found in a different context. A single treatment also tells us very little about behavioural changes beyond the experiment itself. It could be that presenting subjects with cues about responsibility for road maintenance (see, e.g. James et al., 2016) does indeed change attribution of blame, but at the same time, it may also change subjects' perception of road maintenance beyond the duration of the experiment, by making subjects more attentive to the issue and to cues about responsibility..

The main advantage of experiments, as Chapter X [introduction] discusses, is that they allow testing causal claims. The move to experimental methods in the field then implies it is moving from one extreme practice – studying many variables with just a few observations, as for instance happens in case studies – to another where just one variable is studied. Monocausality may then emerge as a method artefact.

4. Implications for the field

The experimental turn in public management research comes with a number of important implications for the field. First, strong internal validity needs to be coupled with stronger realism. This comes with ethical risks. The field also needs to pay closer attention to repeated experiments and dose-response experiments, and move to multi-arm studies in order to move beyond testing the effect of single discrete interventions, and to test entire theories. This will in turn probably entail a move to formal modelling in the discipline. Finally, it is expected that the experimental turn may change academic publishing.

4.1. Enhancing realism to build credibility

Researchers' preoccupation with internal validity and with establishing empirical credibility may mean a lack of attention to realism and external validity (Peters, Langbein & Roberts,

2015). Field experiments, discussed in Chapter x of this book, and replications (chapter Y) go some way to addressing this concern. Still, this emphasis on internal validity risks becoming, as described by Cartwright (2007), a 'vanity of rigor', where internal validity takes precedence over external validity. The desire to use discrete interventions in experiments means that complex policy interventions are sometimes reduced to just one of their core elements. Treatments and interventions in experimental public management research are not only generally discrete; often they are also artificially recreated. This can potentially lead to very unnatural and artificial experimental treatments. Common examples are forcing a subject to look up information on a website or to read instructions, practices which are very far from the reality on how a person perceives government action (media, hearsay, and also official communication). Indeed, as McDermott (2005) points out, 'subjects may behave one way in the relative freedom of an experiment, where there are no countervailing pressures acting on them, but quite another when acting within the constrained organizational or bureaucratic environments in which they work at their political jobs.' (McDermott, 2005: 40). Thus, in the real world, interventions may work differently. For experimental research, this means their external validity is probably the highest in settings where control over the experimental stimuli is the lowest.

Settings used in vignette experiments likewise suffer from a lack of realism, because they need to summarise complex situations in a single vignette. Like for like comparisons, keeping all confounding factors under control, are incredibly hard to find, and also explains the attraction of conjoint analysis. Let me give an example of a study on a public-private comparison of customer reactions to service failure, where we wanted to compare customers' reaction to service repair efforts in a public and a private setting. It proved to be almost impossible to identify a service that was more or less comparable across the two domains, making it nearly impossible to derive any hard evidence about the effects of the treatment (Thomassen et al.,

mimeo). Sometimes, treatments may be so obvious or unrealistic, that the experiment can no longer be considered blind to the subjects, or that subjects lose their interest in the experiment. Using manipulation checks testing for perceived realism during the experiment then is no more than a post-hoc justification for the researcher.

The challenge to researchers then is this: Do you keep experiments as context-free as possible, by testing abstract interventions, or do you try to maximise the experimental realism, thereby introducing contextual and confounding factors that put internal validity at risk. Dickson (2011) contrasts the approaches in economics with highly stylised, abstract experiments, to more naturalistic, context-rich approaches in psychology. Field experiments are a special case, because their setting makes it almost impossible to isolate them from their context. It is for this reason - 'the misunderstanding of exogeneity' that Angus Deaton claims that 'experiments have no special ability to produce more credible knowledge than other methods' (Deaton, 2010: 424).

Increasing realism also means using study participants that resemble those that would normally receive the stimulus. Public management students or paid online workers may react differently to stimuli, just as a study on performance pay in a call centre may tell us little about how subjects in a tax administration would react. Recently, public management scholars have tried to deal with this problem by moving away from student samples and by relying either on representative subject pools, or by doing field experiments involving the population that would normally be subjected to such policy change. High internal validity comes at the cost of narrowness of scope (Cartwright, 2007: 12). Increasing realism requires realism of the experimental setting and the intervention, but also realism with regard to the subject pool.

4.2. The ethics of realism

Increasing the realism of experiments means moving the research closer to the real world. This comes with all kinds of risks. Lijphart, in his seminal 1971 paper 'Comparative politics and the comparative method' described experimental methods as 'the most nearly ideal method for scientific explanation' (Lijphart, 1971: 684), but added that for practical and ethical reasons they can only rarely be used. Many topics and especially treatments are ethically out of bounds. Some of these may still be studied through case studies, or perhaps one could use perceptual measures and self-reports, but running an experiment would be unacceptable. A notable recent example is the Montana voting information experiment, whereby real voters received information on how candidates are placed on an ideological scale (Willis, 2014). In this way researchers intervene in real-world processes and become political actors. In public management, ethical issues proliferate once the research goes hand in hand with policy experiments and field experiments. But even in laboratory or online experiments, ethical issues exist, for instance when one decides to present made-up performance information to subjects, thereby potentially changing their attitude towards government when not debriefed after the experiment.

4.3. A need to repeat, dose, and combine

As I have argued in this chapter, experiments are suited to test the effect of discrete interventions. In order to test theories, however, single experiments are insufficient. Multiple discrete intervention studies tell us more than a one-shot study. Experimentalists in our discipline need to build a tradition of repeated experimentation, as Chapter X on replication argues. To increase internal and external validity, both the subject pools studied and the phenomena studied need to be expanded. An example of the first would be when political scientists manipulate the many types of information they provide to subjects to measure changes in voting intentions. An example of the second would be where the presentation of

information happens in both a private and more social setting, where the influence of peers may play a role in the formation of someone's attitude or judgment.

First, treatments need to be tested in different subject pools, different research settings and different contexts. For example, we need to know whether an experiment on, for instance, the interpretation of absolute versus relative performance information about hospitals done with MPA students can be replicated with, say, parents and school performance information, or with medical doctors and hospital metrics in a context of budget cuts. Only then, we will be able to say anything meaningful about causal mechanisms in a generalised way. A second extension of current experimental work is towards a tradition of dose-response studies, as is common in medical research. One needs to experiment with many different doses to study the effects of a treatment, and thereby not just look at differences in means between two groups, but also at trends, non-linear effects, thresholds, outliers,, and subgroups. Only then, can validity be established.

It is only when our experiments are set up in such a way that they test all steps in a causal mechanism, and all aspects of an intervention in a complex set of trials, involving many different subject pools that they become useful (again, this does not make experiments entirely different from other methods). In other words, the experiment is able to find whether an intervention works, but can only say how and why it works after a series of experiments. Such multi-arm studies are still incredibly rare in the social sciences. This makes doing experiments in public management a more complex and time-consuming way of doing research than it might at first appear, especially when one wants to use experiments to test the effect of real policy interventions in complex environments.

4.4. A move to formal modelling

Just like survey research, experiments are especially useful for exploring topics that have already received attention, where theory development has already started, and where exploratory and descriptive research has taken place. Incidentally, unexpected experimental findings may give rise to new theory development. The deductive nature of experiments requires theories that can be turned into testable hypotheses. This means progress in experimental research may lead to more of an emphasis on formal modelling in the discipline, as Chapter X discusses. Through formal models, theories can be transformed into theoretical frameworks consisting of short, clear and testable statements. In political science, much of the experimental work has focused on elections and voting, a subfield in which formal models have become very important (McDermott, 2005: 50).

Often, however, experiments tend to be treatment based, rather than theory-based. This is especially the case in the RCT tradition where the main function of experiments is to see whether policy interventions work, rather than to test a theory. Not using theory, and just letting the data speak for itself, has limitations. Theories help protect the researcher against confirmation bias, especially when many dependent variables are measured, allowing for selective shopping (an approached that could, however, be tackled using pre-defined analysis plans).

Moving to formal modelling will enable the discipline to test theories, through relying on a successive set of tests, testing elements of that theory, rather than the entire theory. A theory driven-approach will also result in a more focused test of mediation effects to uncover at least part of the mechanisms, and thereby overcoming the critique of experiments being mainly suited to testing what works, not how it works.

4.5. A change in academic publishing

A final implication of the experimental turn is that it will require a different publication culture. Currently, articles in public management are rather long, and contain extensive literature reviews and theoretical elaboration. Transposed to the current experimental turn, this means experimental articles in the discipline tend to be quite lengthy, with the empirical section taking up just about half or less of the entire article length. Often, articles also tend to consist of just as single experiment, or in rare cases include a replication on a different subject pool (see, e.g., Jilke, Van Ryzin & Van de Walle, 2016).

A publication culture more fitting the experimental tradition would mean that theory is moved to papers designing formal models, and that literature review sections will mainly concentrate on reviewing earlier experimental findings. To properly test theories, there is a need for articles containing a series of experiments, where each experiment either tests part of a theory, varies the dose of the experimental manipulation, or replicates the same experiment on different subject pools. In sum, this will result either in a high volume of rather short articles with one or two experiments, or in longer articles presenting a substantial number of experiments simultaneously.

5. Concluding thoughts

Experimental methods are generally seen as the gold standard for evidence. They probably are, if one looks at internal validity. But they are not the only method, and they are just a stage in the research process, which starts with mere observation and description (Gerring, 2012). Once theories are formulated, they can be tested using experiments. This comes with all kinds of problems, notably external validity. For the field of public management, this means that we may be well on our way to producing credible knowledge about unimportant things.

Experiments do respond however to the field's need to gain empirical credibility, especially at a time when economists increasingly focus on public services in their research.

An unexpected positive effect of the experimental turn may be a rapprochement of the academic study of public management and the world of practice. This is somewhat ironic, because the move to methods and designs that are ever more stringent was partly born out of a discontent with the very strong practice-orientation of the field and a reaction of scholars concerned with a lack of scientific rigour in the field. Within the discipline, a strong practice orientation was often seen as antithetical to using advanced methods. Experiments, now, have the potential to bridge this gap between top scholars, mainly interested in innovation and credible methods, and practitioners, who have discovered experiments as a useful method to test policies and real-life public management interventions. Still, experimentalist should not overclaim what they find, and methodological preferences never ought to steer the choice of research topics.

References

Banerjee, A.V., & Duflo, E. (2012). *Poor economics. Barefoot hedge-fund managers, DIY doctors and the surprising truth about life on less than $1 a day.* London: Penguin Books.

Cartwright, N. (2007). 'Are RCTs the gold standard?' *Biosocieties* 2(1): 11‑20.

Coser, L.A. (1975). 'Presidential address: Two methods in search of a substance', *American Sociological Review* 40(6): 691-700.

Deaton, A. (2010). 'Instruments, randomization, and learning about development', *Journal of Economic Literature* 48(2): 424–455.

Dickson, E.S. (2011). Economics vs. psychology experiments: Stylization, incentives, and deception. In: Druckman, J.N., Green, D.P., Kuklinski, J.H., and Lupia, A. (eds). Cambridge handbook of experimental political science. Cambridge: Cambridge University Press, pp. 58-70.

Gadellaa, S., Curry, D., & Van de Walle, S. (2015). 'Hoe bestuurskundig is de bestuurskunde? Nederlandse bestuurskundigen vergeleken met hun Europese vakgenoten', *Bestuurskunde* 25(3): 67-79.

Gerring, J. (2012). 'Mere description', *British Journal of Political Science*, 42(4): 721-746.

Gouldner, A. (1964). *Patterns of Industrial Bureaucracy*. Free Press, New York.

James, O. (2010). 'Performance measures and democracy: Information effects on citizens in field and laboratory experiments', *Journal of Public Administration Research and Theory* 21(3): 399-418.

James, O., Jilke, S., Petersen, C., & Van de Walle, S. (2016). Citizens' Blame of Politicians for Public Service Failure: Experimental Evidence about Blame Reduction through Delegation and Contracting, *Public Administration Review*.

Jilke, S., Van Ryzin, G., & Van de Walle, S. (2016). Responses to decline in marketized public services: An experimental evaluation of choice-overload, *Journal of Public Administration Research and Theory. Online first.*

Kaplan, A. (1998(1964)). *The conduct of inquiry: Methodology for behavioural science*. New Brunswick: Transaction Publishers.

Kaufman, H. (1960). *The forest ranger: A study in administrative behavior*. Baltimore: Johns Hopkins University Press.

Kinder, D.R. (2011). ‘Campbell’s ghost’, in Druckman, J.N., Green, D.P., Kuklinski, J.H., and Lupia, A. (eds). Cambridge handbook of experimental political science. Cambridge: Cambridge University Press, pp. 525-530.

Kinder, D.R., & Palfrey, T.R. (1993). *Experimental foundations of political science*. Ann Arbor: University of Michigan Press.

Lijphart, A. (1971). ‘Comparative politics and the comparative method’, *American Political Science Review* 65(3): 682-693.

Lum, C., & Yang, S.M. (2005). ‘Why do evaluation researchers in crime and justice choose non-experimental methods?’ *Journal of Experimental Criminology* 1(2): 191–213.

McDermott, R. (2005). ‘Experimental methods in political science’, *Annual Review of Political Science* 5: 31-61.

Peters, J., Langbein, J., & Roberts, G. (2015). ‘Policy evaluation, randomize controlled trials, and external validity: A systematic review’, *Ruhr Economic Papers no. 589*. Bochum: Ruhr-Universität Bochum, Department of Economics.

Pawson, R., & Tilley, N. (1997). *Realistic evaluation.* London: Sage.

Sunstein, C.R. (2001). Foreword: ‘On academic fads and fashions’, *Michigan Law Review* 99(6): 1251-1264.

Thomassen, J.-P., Leliveld, M., Ahaus, K., & Van de Walle, S. (2015). *Prosocial compensation after service guarantee violation in for-profit and public settings*. Mimeo.

Willis, D. (28 October 2014). Professors’ research project stirs political outrage in Montana. *New York Times*, retreived from http://www.nytimes.com/2014/10/29/upshot/professors-research-project-stirs-political-outrage-in-montana.html.